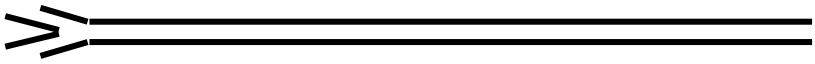
Chapter 3



Quality Control Requirements

3.1 Introduction

As the foundation for method flexibility, EPA will designate an approved method as the "reference method" for each combination of analyte and determinative technique. Any newly developed method that contains a unique combination of analyte and determinative technique would be considered a new method and, when approved, could be designated as the reference method for that unique combination of analyte and determinative technique. Any approved method not designated as a reference method will be designated as an "other approved method." All methods must contain standardized quality control (QC) tests. All reference methods must contain standardized QC tests and specify QC acceptance criteria for each test. The QC acceptance criteria of the reference method must be met when using other approved methods or method modifications. The QC acceptance criteria in the reference method are the performance measures for demonstrating equivalency of method modifications.

The person or organization that develops a reference method for a particular combination of analyte and determinative technique will be responsible for validating the method and for developing the QC acceptance criteria.. QC acceptance criteria will be based on data generated during the method validation study. Under the streamlining initiative, EPA is proposing to require a method validation study that reflects the level of intended use for a method. This three-tiered approach to method validation is explained in Chapter 4. EPA believes that the tiered approach will minimize the validation requirements of limited-use methods (single-laboratory and single-industry use) and will focus resources on validation of methods that are intended for nationwide use. Because QC acceptance criteria will be developed from validation studies and because the validation requirements vary with each tier, the statistical procedures used to develop the criteria will vary by tier.

Some methods presently approved at 40 CFR parts 136 and 141 do not contain acceptance criteria for all standardized QC tests. In the streamlining proposal, EPA has provided supplementary QC acceptance criteria for methods proposed as reference methods that do not already contain QC acceptance criteria. QC acceptance criteria must be developed for and specified in all new methods that will be approved as reference methods.

This chapter describes the three method validation tiers, lists and describes the standardized QC tests required in all approved methods, and outlines procedures for developing QC acceptance criteria for new methods at Tiers 1, 2, and 3. The key concepts presented and discussed in this chapter are: *standardized QC tests, calibration linearity, calibration verification, absolute and relative retention time precision, initial precision and recovery, ongoing precision and recovery, analysis of blanks, surrogate or labeled compound recovery, matrix spike and matrix spike duplicate, method detection limit demonstration, reference sample analysis*, and *QC acceptance criteria*.

3.2 Description of Tiers

Tier 1 refers to new methods or method modifications that will be used by a single laboratory for one or more matrix type(s). As used in streamlining, a matrix type is a sample medium (e.g., air, water, soil) with common characteristics across a given industrial category or subcategory. Validation requirements for Tier 1 reflect this limited use and correspondingly require single-laboratory testing in the matrix type(s) in which the method will be used. In response to comments received during public meetings, EPA has refined requirements for this tier to allow single laboratories to apply new or modified methods to an unlimited number of matrix types after the method has been validated in nine discrete matrix types. If results of Tier 1-Multiple Matrix Type validation studies are to be applied to a different medium, each medium must be represented in the samples tested in the validation study. Procedures for developing QC acceptance criteria for Tier 1 methods are given in Section 3.4.1.

Tier 2 refers to new methods or method modifications that will be used by multiple laboratories analyzing samples of one matrix type from a single industrial category or subcategory. Validation at Tier 2 requires a three-laboratory interlaboratory study in the matrix type(s) in which the method will be used. Procedures for developing QC acceptance criteria for Tier 2 methods are given in Section 3.4.2.

Tier 3 refers to new methods or method modifications that will be used on a nationwide basis by all laboratories for all matrix types. Validation at Tier 3 requires a nine-laboratory interlaboratory study on nine matrix types. Validation must be performed on a minimum of nine matrix types in each sample medium to which the method will be applied. Procedures for developing QC acceptance criteria for Tier 3 methods are given in Section 3.4.3.

3.3 Standardized Quality Control Tests

Under this initiative, standardized QC tests are required for use with currently approved methods and are a mandatory component of all new methods. The standardized QC tests are as follows:

- calibration linearity
- calibration verification
- absolute and relative retention time precision (for chromatographic analyses)
- initial precision and recovery
- ongoing precision and recovery
- analysis of blanks
- surrogate or labeled compound recovery
- matrix spike and matrix spike duplicate precision and recovery (for non-isotope dilution analyses)
- method detection limit demonstration
- analysis of a reference sample

These tests are described in Sections 3.3.1 - 3.3.10 below.

3.3.1 Calibration Linearity

The calibration linearity specification establishes a break point between a straight line through the origin and a straight line not through the origin or a curved calibration line. This break point is specified as a maximum relative standard deviation (RSD=100s/$\overline{X}$, expressed as percent) of the:

- relative response (RR) for isotope dilution calibration,
- response factor (RF) for internal standard calibration, or
- calibration factor (CF) for external standard calibration,

below which an averaged RR, RF, or CF may be used. The number of calibration points is dependent on the error of the measuring technique. Measurement technique error is determined by (1) calibrating the instrument at the minimum level (ML) of quantitation and a minimum of two additional points, and (2) determining the RSD of the RR, RF, or CF. For most analyses, such as the determination of semi-volatile organic compounds by extraction, concentration, and gas chromatography, the measuring instrument is calibrated, and sample preparation processes are excluded from the calibration process; for others, such as the determination of purgeable organic compounds by purge-and-trap gas chromatography, calibration encompasses the entire analytical process. Table 3-1 below gives the number of calibration points required depending on the calibration linearity.

Table 3-1: Minimum Number of Points Required for Calibration[1]

Percent RSD[2]	Minimum Number of Calibration Points
0 - <2	1[3]
2 - <10	3
10 - <25	5
>25	7

[1] Based on Rushneck et al. 1987. Effect of number of calibration points on precision and accuracy of GC/MS, in Proceedings of Tenth Annual Analytical Symposium, USEPA: Washington, DC.

[2] Percent RSD shall be determined from the calibration linearity test.

[3] Assumes linearity through the origin (0,0). For analytes for which there is no origin (such as pH), a two-point calibration shall be performed.

The ideal calibration is a straight line that intersects the origin (zeroth order). In practice, no calibration line constructed from three or more calibration points will intersect the exact origin (0.000 ..., 0.000 ...). If, however, an error band is constructed around the calibration line, the error band will include the origin for most calibrations. The use of an averaged RR, RF, or CF is an attempt to represent the calibration with a single value that includes all of the points, including the origin, within the error represented by the RSD.

The maximum RSD specification is applicable to calibration with three or more calibration points. For some methods, a least-squares regression and correlation coefficient have been used. However, an unweighted least-squares regression that covers a large range will inappropriately weight the highest calibration point(s). Equally weighing each point in a least-squares regression produces the same result as an averaged RR, RF, or CF. Therefore, unless the method specifies use of a least-

squares regression and/or correlation coefficient, the RSD of the RR, RF, or CF must be used to establish calibration linearity.

Calibrations higher than zeroth order calibration (straight line through the origin) are required when the linearity criterion cannot be met. For most instruments and analytical systems, these calibrations are first order (linear not through the origin; $y = mx + b$) and second order ($y = ax^2 + bx + c$). A second or higher order calibration may be justified when an analyte can only be determined with a method that uses a determinative technique with a nonlinear response over the calibration range. A second order or higher order calibration may be used, provided that the calibration increases monotonically. Monotonically means that the response is successively greater at successively higher concentrations. For example, an immunoassay typically requires a third ($y = ax^3 + bx^2 + cx + d$) or fourth ($y = ax^4 + bx^3 + cx^2 + dx + e$) order calibration, although not all of the terms in these equations may be needed.

EPA believes that most instruments and analytical systems are linear over a range large enough to preclude the need for second order or higher calibration. If the linear range of any of these systems is limited, sample dilution and reanalysis should be performed to bring the concentration within the linear range, rather than extend the calibration into a nonlinear region of the instrument response. EPA discourages use of higher than first-order calibration because responses in the nonlinear region of the instrument response can mask curvature in the response that may be attributable to preparation of an inaccurate standard. EPA requires that all calculations of concentrations of analytes in blanks, field samples, QC samples, and samples prepared for other purposes be based on an averaged RR, RF, or CF, or on a calibration curve.

3.3.2 Calibration Verification

This test is used to periodically verify that instrument performance has not changed significantly from calibration. Verification is based on time (e.g., working day; 12-hour shift) or on the number of samples analyzed in a batch (e.g., after every 10th sample). The terms "shift" and "batch" should be specified in the method. If not, the general rule has been that calibration verification is performed every 12-hour shift on instruments used for determination of organic analytes and every 10th sample on instruments used for determination of metals. However, the over-riding rule should be that verification is performed frequently enough to assure that the response of the instrument or analytical system has not drifted significantly from calibration.

Calibration verification tests are typically performed by analyzing a single standard in the concentration range of interest for the target analyte(s). In most methods, this standard is in the range of 1 - 5 times the minimum level (ML) of quantitation and is at the same level as one of the standards used for calibration. The calibration verification standard concentration should be within 1 - 5 times the ML rather than at a "midpoint" concentration because specifying the midpoint can be interpreted as one-half (½) the highest calibration point. Using a concentration this high when the calibration covers orders of magnitude may lead to erroneous results, because this midpoint standard may be far removed from the range where most measurements will be made.

If the calibration is linear through the origin (as defined by linearity criteria in Table 3-1), specifications for calibration verification are developed to define the allowable deviation of the RR, RF, or CF of the calibration verification standard from the averaged RR, RF, or CF of the calibration. If linearity criteria for calibration are not met, specifications for calibration verification are developed

to define the allowable deviation of the RR, RF, or CF of the calibration verification standard from a specific point on the calibration curve.

For calculation of analyte concentrations, the averaged RR, RF, or CF, or the calibration curve is always used; i.e., the calibration is not updated to the RR, RF, CF or the single point verification. Updating the calibration to a single point after establishing an averaged RR, RF, or CF, or a calibration curve is equivalent to performing a single-point calibration. This updating procedure, which is sometimes termed "continuing calibration," is unacceptable and shall not be used because it nullifies the statistical power of the full calibration.

3.3.3 Absolute and Relative Retention Time Precision

Absolute retention time (RT) and relative retention time (RRT) are the QC criteria used in chromatographic analyses to aid in the identification of each detected analyte and to confirm that sufficient time was allowed for the chromatographic separation of the analytes in complex mixtures. These criteria also prevent laboratories from accelerating the analysis in an effort to reduce costs, only to find that complex mixtures cannot be adequately resolved.

A minimum RT specification is developed for those methods in which a minimum analysis time must be established to ensure separation of the analytes in complex mixtures including known or expected interferences. An RT precision specification is developed for identification of an analyte by external standard measurements, and an RRT precision specification is developed for (1) each analyte relative to its labeled analog by isotope dilution measurements, (2) each labeled compound relative to its internal standard for isotope dilution measurements, and (3) each analyte relative to an internal standard for internal standard measurements.

3.3.4 Initial Precision and Recovery

The initial precision and recovery (IPR) test, also termed a "startup test," is used for initial demonstration of a laboratory's capability to produce results that are at least as precise and accurate as results from practice of the method by other laboratories. The IPR test also is used to demonstrate that a method modification will produce results that are as precise and accurate as results produced by the reference method. The IPR test consists of analyzing at least four replicate aliquots of a reference matrix spiked with the analytes of interest and with either surrogate compounds or, for isotope dilution analysis, labeled compounds. The concentration of the target analytes in the spike solution may vary between one and five times the concentration used to establish the lowest calibration point (e.g., one to five times the ML). The spiked aliquots are carried through the entire analytical process. The IPR test is performed by the laboratory before it utilizes a method or a method modification for analysis of actual field samples. Specifications are developed for the permissible range of recovery for each analyte and for an upper limit on the standard deviation or RSD of recovery.

3.3.5 Ongoing Precision and Recovery

The ongoing precision and recovery (OPR) test, sometimes termed a "laboratory control sample," "quality control check sample," or "laboratory-fortified blank," is used to ensure that the laboratory remains in control during the period that samples are analyzed, and it separates laboratory performance from method performance in the sample matrix. The test consists of a single aliquot of reference matrix spiked with the analyte(s) of interest and carried through the entire analytical process with each batch of samples. Typically, the concentration of the target analyte(s) in the same as the

concentration used in the IPR test. Specifications are developed for the permissible range of recovery for each analyte.

3.3.6 Analysis of Blanks

Blanks are analyzed either periodically or with each sample batch to demonstrate that no contamination is present that would affect the analysis of standards and samples for the analytes of interest. The period or batch size is defined in each method. Typical periods and batch sizes are one per shift or one for every 10 or 20 samples, but more or fewer may be required, depending upon the likelihood of contamination.

For most methods, QC acceptance criteria for blanks are given in each method and are specified as the concentration or amount of the analyte allowed in each type of blank. The source of contamination in a blank must be identified and eliminated before the analysis of standards and samples may begin. Samples analyzed with an associated contaminated blank must be reanalyzed. Methods for which blank contamination cannot be eliminated should specify blank-subtraction procedures.

3.3.7 Surrogate or Labeled Compound Recovery

The surrogate or labeled compound recovery is used to assess the performance of the method on each sample. For this test, surrogates or stable, isotopically labeled analogs of the analytes of interest are spiked into the sample and the recovery is calculated. Specifications are developed for the permissible range of recovery for each surrogate and/or labeled compound from each sample.

3.3.8 Matrix Spike and Matrix Spike Duplicate

The matrix spike and matrix spike duplicate (MS/MSD) test is used in non-isotope dilution methods to assess method performance in the sample matrix. In most cases, analytes of interest are added to a field sample aliquot that is then thoroughly homogenized and split into two spiked replicate aliquots.[1] One of these replicates is identified as the matrix spike sample and the other is identified as the matrix spike duplicate sample. The recoveries of the analytes, relative to the spike, are determined in each sample. The precision of the determinations also is assessed by measuring the relative percent difference (RPD) between the analyte concentrations measured in the MS and MSD. The MS and MSD samples should each be spiked at a level that results in the concentration of the target analyte(s) being

- At the regulatory compliance limit or
- One to five times the background concentration of unspiked field sample, or
- At the level specified in the method, whichever is greater.

If the background concentration in the field sample is so high that the spike will cause the calibration range of the analytical system to be exceeded, the sample is spiked after the field sample is diluted by the minimal amount necessary for this exceedance not to occur. This dilution of the sample to stay within the calibration range of the analytical system for the target analyte is necessary to verify that

[1] For analytes, such as oil and grease, that adhere to container walls and cannot be adequately homogenized, it is not possible to divide a spiked aliquot into two replicate aliquots. In these cases, two field samples are collected and each field sample is spiked with identical concentrations of the analytes of interest to produce an MS and MSD sample.

the sample matrix has not prevented reliable determination of the analyte. Specifications are developed for the permissible range of recovery and RPD for each analyte.

3.3.9 Demonstration of Method Detection Limit

Nearly all of the 40 *CFR* part 136, Appendix A methods contain method detection limits (MDLs), although few of the methods explicitly require laboratories to demonstrate their ability to achieve these MDLs. Under the streamlining initiative, EPA will develop MDLs for each analyte in each existing reference method, and organizations developing new reference methods will be required to develop analyte-specific MDLs applicable to those methods. The MDLs published for each reference method will be used as an indicator of method performance. Each laboratory that intends to practice a method will be required to demonstrate achievement of an MDL that meets the criteria specified in the reference method. The MDL must be determined according to the procedures specified at 40 CFR part 136, Appendix B. The Appendix B MDL calculation and analytical procedure is described in Section 3.4.1.1.

3.3.10 Reference Sample Analysis

The most common reference sample is a Standard Reference Material (SRM) from the National Institute of Standards and Technology (NIST). The reference sample and the period for its use are specified in each method. EPA is considering setting acceptance criteria for standard reference materials to be within some percentage of the stated value based on the variability of measurement for that analyte. One possible indicator of that variability is the relative standard deviation calculation for the initial precision and recovery samples. Corrective action to be taken when the acceptance criteria are not met should involve identifying the samples affected, determining the amount of the effect, and if the effect is significant, determining the impact of the effect on the environmental samples analyzed.

3.4 Development of Quality Control Acceptance Criteria

The procedures for developing QC acceptance criteria at Tier 1, Tier 2, and Tier 3 methods are described in Sections 3.4.1, 3.4.2, and 3.4.3, respectively. Under the streamlining initiative, interlaboratory study data are required to develop QC acceptance criteria for Tier 2 and Tier 3 methods. Although these studies are not necessary for Tier 1 methods, interlaboratory study data may be available. If interlaboratory data are available for a Tier 1 method, these data should be used to develop QC acceptance criteria for Tier 1 methods by following the Tier 2 or Tier 3 procedures described in Section 3.4.2 or 3.4.3, respectively. Where possible, interlaboratory study data used for development of QC acceptance criteria should be derived from study designs that follow the basic principles outlined in *Guidelines for Collaborative Study Procedures to Validate Characteristics of a Method of Analysis*, JAOAC *72* No. 4, 1989, *Use of Statistics to Develop and Evaluate Analytical Methods* (published by AOAC-International), ASTM Standard D-2777 (published by ASTM), or other well-established and documented principles.

The statistical procedures described in Sections 3.4.1 and 3.4.2 for Tier 1 and Tier 2 are based on the use of interlaboratory multipliers. These multipliers were derived from a comparison of

intralaboratory versus interlaboratory variability in the development of EPA Method 1625.[2] The variation in the interlaboratory multiplier used is directly related to the number of laboratories used at each of the two tiers. The general relationship follows the concept that an increase in the number of laboratories used results in a decrease in the interlaboratory multiplier.

If the method being developed is applicable to a large number of compounds, the organization responsible for developing QC acceptance criteria for the method may wish to consider the use of statistical allowances for simultaneous compound testing. Procedures associated with simultaneous compound testing and the develoment of applicable QC acceptance criteria can be found at 49 FR 43242 and in the Method 1625 validation study report.[3]

3.4.1 Quality Control Acceptance Criteria Development for New Methods at Tier 1

Method validation at Tier 1 consists of (1) using the new method to perform an MDL study in accordance with the MDL procedure described at 40 *CFR* part 136, Appendix B, (2) using the results of this MDL study to establish an ML, and (3) running, in a single laboratory, a test of four spiked reference matrix samples and four spiked samples of the sample matrix (or matrices) to which the method is to be applied. The spike level of these reference matrix and real-world matrix IPR samples must be in the range of one to five times the ML, or at the regulatory compliance level, whichever is higher.

3.4.1.1 Method detection limit and minimum level

An MDL must be determined for each target analyte using the procedure detailed at 40 *CFR* part 136, Appendix B. This procedure involves spiking seven replicate aliquots of reference matrix or the sample matrix with the analytes of interest at a concentration within one to five times the estimated MDL. The seven aliquots are then carried through the entire analytical process, and the standard deviation of the seven replicate determinations is calculated. The standard deviation is multiplied by 3.14 (the Student's *t* value at 6 degrees of freedom) to form the MDL. If the spike level is greater than five times the determined MDL, the spike level must be reduced and the test repeated until the MDL is within a factor of five of the spike level. The precautions concerning blanks and the effect of the matrix, and the detailed steps in 40 *CFR* part 136, Appendix B must be observed to arrive at a reliable MDL. In addition, if the analytical system or instrument fails to produce a positive response for any of the seven replicates (i.e., produces a zero or negative result), the MDL procedure must be repeated at a higher spike level.

The ML is established by multiplying the MDL by 3.18 and rounding to the number nearest to (1, 2, or 5) x 10^n, where n is positive or negative integer. The purpose of rounding is to allow instrument calibration at a concentration equivalent to the ML without the use of unwieldy numbers. The use of 3.18 results in an overall standard deviation multiplier of 10, which is consistent with the American Chemical Society's (ACS) limit of quantitation (LOQ) (P. S. Porter et al., *Environ. Sci. Technol.*, 22, 1988).

---

[2] Appendix I, "Estimation of Variance Components", of the *Interlaboratory Validation of U.S. Environmental Protection Agency Method 1625A*, available from the EPA Sample Control Center operated by DynCorp, Alexandria, VA 22314, 703/519-1140.

[3] *Interlaboratory Validation of U.S. Environmental Protection Agency Method 1625A*. See above.

Once established, the ML is used as the lowest calibration point. The instrument or analytical system is then calibrated at the ML and a minimum of two additional points to assess calibration linearity (Section 3.4.1.2) and to determine the number of calibration points required and how these points are spaced (Section 3.3.1).

3.4.1.2 Calibration linearity

Establish the RSD of the response factors (RFs), calibration factors (CFs), or relative responses (RRs) based on the precision of the determinative technique, as described in Section 3.3.1, and as determined in Section 3.4.1.1. If the RSD is < 2%, a one- or two-point calibration is employed (see Section 3.1.1) and it is unnecessary to establish a limit for calibration linearity.

If three or more calibration points are required, the RSD for the RFs, CFs, or RRs is determined as follows:

(1) Determine the average response factor ($\overline{\overline{RF}}$), calibration factor ($\overline{\overline{CF}}$), or relative response ($\overline{\overline{RF}}$) for each analyte from the initial calibration:

$$\overline{\overline{RF}} = (RF_1 + RF_2 + ... + RF_n)/n$$

where n is the number of calibration points.

(2) Determine the RSD using $\overline{RF}$, $\overline{CF}$,or $\overline{RF}$ and the standard deviation (s) of the RF ,CF, or RR for each analyte from the initial calibration. The RSD is determined by:

$$RSD = 100s/(\overline{\overline{RF}})$$

(3) Develop a maximum RSD as follows:

$$RSD_{max} = kRSD$$

where k is the square root of the 95th percentile of an F distribution with degrees of freedom corresponding to the number of points in the initial calibration minus 1 in both numerator and denominator. For a three point calibration, the value of k is 4.4, and for a five-point calibration, the value of k is 2.5.

Note: In the above equations, the $\overline{\overline{RF}}$ and RF terms should be replaced by $\overline{CF}$ and CF or $\overline{RR}$ and RR terms where appropriate.

3.4.1.3 Calibration verification

Using the average response factor ($\overline{\overline{RF}}$), calibration factor ($\overline{\overline{CF}}$), or relative response ($\overline{\overline{RR}}$) from the initial calibration, calculate the upper and lower QC acceptance criteria for the calibration verification as follows:

(1) Calculate a multiplier, k, as the 97.5th percentile of a Student's *t* distribution with n - 1 degrees of freedom times the square root of (1 + 1/n), where there are n points in the calibration. For a three point calibration, the n - 1 Student's *t* value is 4.3, and for a five point

calibration, the Student's *t* value is 2.8, resulting in values for k of 5.0 for a three point and 3.0 for a five point calibration.

(2) Calculate the upper and lower QC acceptance criteria for the response or calibration factors for each analyte by developing a window around the average response factor found in the initial calibration by:

$$\text{Lower limit} = \overline{\overline{RF}} - ks$$
$$\text{Upper limit} = \overline{\overline{RF}} + ks$$

where k is the multiplier determined in Step 1 and s is the standard deviation determined in 3.4.1.2, Step 2.

Note: In the above equations, the $\overline{\overline{RF}}$ terms should be replaced by $\overline{\overline{CF}}$ or $\overline{\overline{RR}}$ terms where appropriate.

3.4.1.4 Initial and ongoing precision and recovery

For Tier 1 methods, an IPR test must be performed in both a reference matrix (usually, reagent water) and the sample matrix of interest. Results of the reference matrix IPR tests are used to generate QC acceptance criteria for IPR and OPR tests as described in this subsection. Results of the sample matrix IPR test are used to develop QC acceptance criteria for the MS/MSD tests (see Section 3.4.1.5 below). The reference matrix IPR test is performed by analyzing four aliquots of the reference matrix spiked with the target analyte(s) at the concentration determined in Section 3.3.4.

Calculate the QC acceptance criteria for the IPR and OPR tests using results of the test of the reference matrix per the following steps:

(1) Calculate the average percent recovery ($\overline{X}$), the standard deviation of recovery (s), and the relative standard deviation (RSD=100s/$\overline{X}$) of the four IPR results.

(2) IPR QC acceptance criterion for precision - To approximate a 95% confidence interval for precision, the RSD is multiplied by the square root of the 95th percentile of an F distribution with 3 degrees of freedom in the numerator and denominator. The resulting multiplier on the RSD will then be 3.0. The QC acceptance criterion for precision in the IPR test (RSD_{max}) is calculated as follows:

$$RSD_{max} = 3.0RSD.$$

(3) IPR QC acceptance criteria for recovery - Calculate the QC acceptance criteria for recovery in the IPR test by constructing a ± 5.3s window around the average percent recovery ($\overline{X}$). This factor comes from the 97.5th t percentile for 3 degrees of freedom, multiplied by $\sqrt{1.15(1+1)+(1/4+1/4)}$ to account for interlaboratory variability and the estimation of the mean:

$$\text{Lower limit (\%)} = \overline{X} - 5.3s$$
$$\text{Upper limit (\%)} = \overline{X} + 5.3s$$

(Based on EPA's interlaboratory validation study of Method 1625, the additional variance due to interlaboratory variability is estimated as $1.15s^2$.)

(4) OPR QC acceptance criteria for recovery - A similar miltiplier is used as for the IPR test but the second factor is $\sqrt{1.15(1+1) + (1 + 1/4)}$, so the multiplier is 6.0. Calculate the QC acceptance criteria for recovery in the OPR test by constructing a ± 6.0s window around the average percent recovery $\overline{\overline{X}}$:

$$\text{Lower limit (\%)} = \overline{X} - 6.0s$$
$$\text{Upper limit (\%)} = \overline{X} + 6.0s$$

Note: For highly variable methods, it is possible that the lower limit for recovery for both the IPR and OPR analyses will be a negative number. In these instances, the data should either be log-transformed and the recovery window recalculated, or the lower limit established as "detected," as was done with some of the 40 CFR part 136, Appendix A methods (49 FR 43234).

3.4.1.5 Matrix spike and matrix spike duplicate

As noted above, an IPR test must be performed in both an appropriate reference matrix and the sample matrix of interest for Tier 1 new methods. The results of the sample matrix IPR test are used to develop acceptance criteria MS/MSD analyses. Sample matrix IPR tests are performed by: (1) determining the background concentration of the sample matrix, (2) spiking four replicate aliquots of the sample matrix at a concentration equal to the regulatory compliance limit, one to five times the ML determined in Section 3.4.1.1, or one to five times the background concentration of the sample, whichever is greater, and (3) analyzing each of these spiked replicate samples.

Calculate the QC acceptance criteria for the recovery of MS and MSD samples as follows:

(1) Calculate the average percent recovery ($\overline{X}$) and the standard deviation of recovery (s) of each target analyte in the sample matrix IPR aliquots.

(2) Calculate the QC acceptance criteria for recovery in the MS and MSD tests by constructing a ± 6.0s window around the average percent recovery ($\overline{X}$) (derived the same as for the OPR test above):

$$\text{Lower limit (\%)} = \overline{X} - 6.0s$$
$$\text{Upper limit (\%)} = \overline{X} + 6.0s$$

Note: For highly variable methods, it is possible that the lower limit for recovery for both IPR and OPR analysis will be a negative number. In these instances, the data should either be log-transformed and the recovery window recalculated, or the lower limit established as "detected," as was done with some of the 40 CFR part 136, Appendix A methods (49 FR 43234).

Calculate the QC acceptance criteria for the relative percent difference between the MS and MSD as follows:

(1) Calculate the relative standard deviation (RSD) of the recoveries of each target analyte in the sample matrix IPR aliquots as follows:

$$RSD = 100s/\overline{X}$$

(2) Calculate the relative percent difference (RPD) criterion as follows:

$$RPD_{max} = 4.5RSD$$

This multiplier is calculated as $\sqrt{2}$ times the square root of the 95th percentile of an F distribution with 1 and 3 degrees of freedom.

3.4.1.6 Absolute and relative retention time

Determine the average retention time, $\overline{\overline{RT}}$ (and/or average relative retention time, $\overline{\overline{RRT}}$), and the standard deviation (s) for each analyte and standard. Determine the upper and lower retention time (or relative retention time) limits using the following:

$$\text{Lower limit} = \overline{\overline{RT}} - ts\sqrt{1 + \frac{1}{n}}$$

$$\text{Upper limit} = \overline{\overline{RT}} + ts\sqrt{1 + \frac{1}{n}}$$

The relative retention time upper and lower limits are determined by replacing $\overline{\overline{RT}}$ with $\overline{\overline{RRT}}$ in the equations above. The t value is the 97.5th percentile of a t distribution with n - 1 degrees of freedom, where n is the number of retention time or relative retention time values used.

3.4.1.7 Blanks

Establish the QC acceptance criteria for blanks. The usual requirement is that the concentration of an analyte in a blank must be below the ML or below one-third (1/3) the regulatory compliance level, whichever is higher. In instances where the level of the blank is close to the regulatory compliance level or the level at which measurements are to be made, it may be necessary to require multiple blank measurements and establish the QC acceptance criteria based on the average of the blank measurements plus two standard deviations of the blank measurements.

3.4.1.8 Reference sample

Establish the QC acceptance criteria for the reference sample based on the error provided with the reference sample.

3.4.2 Quality Control Acceptance Criteria Development for New Methods at Tier 2

Method validation at Tier 2 consists of running tests on a single matrix type collected from three different facilities in the same industrial subcategory, with the sample being analyzed in three separate laboratories (see 40 *CFR* parts 405 - 503 for industrial categories and subcategories). If the matrix type being validated is drinking water, then tests shall be run on a drinking water matrix collected from three different sources or on three drinking water samples that each have different characteristics (see Section 4.4.2).

Each of the three laboratories will need to run a full suite of tests, beginning with an MDL study to determine the appropriate ML, followed by calibration, IPR, OPR, and blank analyses, along

with a pair of MS/MSD analyses for each sample matrix. Results from each laboratory will be submitted to the organization responsible for developing the method. That organization will use the laboratory results to develop QC acceptance criteria as described in the following subsections.

3.4.2.1 Method detection limit and minimum level

Each laboratory participating in the MDL study must perform an MDL test as described in Sections 3.4.1.1 and 6.3.2.9. The organization responsible for developing the new method must establish an MDL for the method, using a pooled MDL from the three laboratories. The precautions concerning blanks and the effect of the matrix, and the detailed steps in 40 *CFR* part 136, Appendix B must be observed to arrive at a reliable MDL.

A pooled MDL is calculated from m individual laboratory MDLs by comparing the square root of the mean of the squares of the individual MDLs and multiplying the result by a ratio of t-values to adjust for the increased degrees of freedom.

$$MDL_{pooled} = \sqrt{\frac{d_1\left(\frac{MDL_{(Lab\,1)}}{t_{(0.99,d_1)}}\right)^2 + d_2\left(\frac{MDL_{(Lab\,2)}}{t_{(0.99,d_2)}}\right)^2 + \ldots d_m\left(\frac{MDL_{(Lab\,m)}}{t_{(0.99,d_m)}}\right)^2}{d_1 + d_2 + \ldots d_m}}\; t_{(0.99,d_1+d_2+\ldots d_m)} \quad ,$$

where m = the number of laboratories, and d_i = the number of replicates used by lab i to derive the MDL. In the case of 3 laboratories with 7 replicates per laboratory, the equation simplifies to:

$$MDL_{pooled} = \sqrt{\frac{MDL^2_{(Lab\,1)} + MDL^2_{(Lab\,2)} + MDL^2_{(Lab\,3)}}{3}}\;\frac{2.55}{3.14}$$

The organization responsible for developing the method also must use this pooled MDL to develop an ML. Procedures for determining the ML are given in Section 3.4.1.1. Once established, the ML is used as the lowest calibration point. The instrument or analytical system is then calibrated at the ML and a minimum of two additional points to assess calibration linearity (Section 3.4.1.2) and to determine the future number of calibration points required and how these points are spaced (Section 3.3.1).

3.4.2.2 Calibration linearity

Establish the RSD of the response factors (RFs), calibration factors (CFs),or relative responses ($\overline{\mathbf{RR}}$s) based on the precision of the determinative technique, as described in Section 3.3.1 and as determined in Section 3.4.2.1. If the RSD is < 2%, a one- or two-point calibration is employed (see Section 3.1.1) and it is unnecessary to establish a limit for calibration linearity.

If three or more calibration points are required, the upper limit on the RSD of the RFs or CFs is determined as follows:

(1) Calculate the overall average RF ($\overline{\overline{RF}}$), overall average CF ($\overline{\overline{CF}}$), or overall average RR ($\overline{\overline{RR}}$) for each analyte using the individual results from all three laboratories. For example, for a 3-point calibration using RFs:

$$\overline{\overline{RF}} = (RF_{1(lab\ 1)} + RF_{2(lab\ 1)} + RF_{3(lab\ 1)} + RF_{1(lab\ 2)} + RF_{2(lab\ 2)} + RF_{3(lab\ 2)} + RF_{1(lab\ 3)} + RF_{2(lab\ 3)} + RF_{3(lab\ 3)})/9$$

(2) Calculate the pooled within-laboratory standard deviation (s_w) of the RF, CF, or RR for each analyte from all three laboratories. The pooled within-laboratory standard deviation is calculated as the square root of the mean of the squares of the sample standard deviations of the calibration results at each individual laboratory.

$$s_w = \sqrt{\frac{s^2_{(lab\ 1)} + s^2_{(lab\ 2)} + s^2_{(lab\ 3)}}{3}}$$

(3) Calculate the relative standard deviation of the RF, CF, or RR for each analyte as:

$$RSD = \frac{100 s_w}{\overline{\overline{RF}}}$$

(4) Calculate the maximum RSD of the RF, CF, or RR for each analyte as follows:

$$RSD_{max} = kRSD$$

where k is the square root of the 95th percentile of an F distribution with n - 1 degrees of freedom in the numerator and m(n - 1) degrees of freedom in the denominator, where m is the number of laboratories and n is the number of calibration points. For three laboratories using a three point calibration, (m=3, n = 3), the value of k is 2.3, and for three laboratories using a five point calibration (m=3, n = 5), the value of k is 1.8.

Note: In the above equations, the $\overline{\overline{RF}}$ and RF terms should be replaced by $\overline{\overline{CF}}$ and CF or $\overline{\overline{RR}}$ and RR terms where appropriate.

3.4.2.3 Calibration verification

Using the average response factor, calibration factor, or relative response from the initial calibration, calculate the upper and lower QC acceptance criteria for calibration verification as follows:

(1) Determine "k" by multiplying the 97.5th percentile of a Student's *t* distribution with m(n-1) degrees of freedom times the square root of (1 + 1/mn), where there are n points in the calibration and m laboratories:

$$k = t\sqrt{(1 + \frac{1}{mn})}$$

For a three point calibration with three laboratories, the m(n - 1) Student's *t* value is 2.4, and for a five point calibration, the Student's *t* value is 2.2, resulting in combined multipliers of 2.5 for a three point calibration, and 2.3 for a five point calibration.

Multiply k by the pooled standard deviation, s_w ,found in Section 3.4.2.2.

(2) Calculate the upper and lower QC acceptance criteria for the response factors, calibration factors, or relative responses for each analyte by developing a window around the average response factor, calibration factor, or relative response by:

$$\text{Lower limit} = \overline{RF} - ks_w$$
$$\text{Upper limit} = \overline{RF} + ks_w$$

Note: In the above equations, the $\overline{RF}$ terms should be replaced by $\overline{CF}$ or $\overline{RR}$ terms where appropriate.

3.4.2.4 Initial and ongoing precision and recovery

For the IPR and OPR tests, QC acceptance criteria are calculated using the average percent recovery and the standard deviation of recovery from the IPR tests on four aliquots of the reference matrix and the OPR test of one aliquot of the reference matrix (for a total of five samples) in the three laboratories, as follows:

(1) Calculate the average percent recovery ($\overline{X}$) for each analyte based on all data points from all laboratories, the between-laboratory standard deviation (s_b) of the mean results for each of the three laboratories (standard deviation of the three lab means $\overline{X}_{(lab\ 1)}$, $\overline{X}_{(lab\ 2)}$, $\overline{X}_{(lab\ 3)}$), and the pooled within-laboratory standard deviation (s_w) of the 5 samples calculated as in 3.4.2.2. Note: the organization responsible for developing the method must ensure that all laboratories are spiking IPR and OPR samples at the same concentration.

(2) IPR QC acceptance criterion for precision - To calculate a 95% confidence interval for precision, the RSD (computed as s_w divided by $\overline{X}$) is multiplied by the square root of a 95th percentile F value with 3 degrees of freedom in the numerator and 4m degrees of freedom in the denominator, where m = the number of laboratories. The resulting multiplier on the RSD for three laboratories will then be 1.9. The QC acceptance criterion for precision in the IPR test (RSD_{max}) is calculated as follows:

$$RSD_{max} = 1.9RSD$$

(3) IPR QC acceptance criteria for recovery -Calculate the combined standard deviation for interlaboratory variability and estimation of the mean (s_c) as:

$$s_c = \sqrt{(1+\frac{1}{m})s_b^2 + (\frac{1}{4}-\frac{1}{n})s_w^2} \quad ,$$

where m = the number of laboratories, and n = the number of data points per laboratory. For 3 laboratories and 5 data points per laboratory,

$$s_c = \sqrt{\frac{4}{3}s_b^2 + \frac{1}{20}s_w^2} \quad .$$

(4) Calculate the QC acceptance criteria for recovery in the IPR test by constructing a ± 3.2 s_c window around the average percent recovery ($\overline{\overline{X}}$, where 3.2 is the 97.5th percentile Student's *t* value for 3 degrees of freedom (an estimated degrees of freedom based on the variance ratios observed with EPA Method 1625):

$$\text{Lower limit}(\%) = \overline{\overline{X}} - 3.2s_c$$
$$\text{Upper limit}(\%) = \overline{\overline{X}} + 3.2s_c$$

If more than 3 laboratories are used, the degrees of freedom for *t* will increase, but a complete calculation is beyond the scope of this document. An approximation of degrees of freedom equal to the number of laboratories will serve for most situations.

(5) OPR QC acceptance criteria for recovery - Calculate the combined standard deviation for interlaboratory variability and estimation of the mean (s_c) as:

$$s_c = \sqrt{(1+\frac{1}{m})s_b^2 + (1-\frac{1}{n})s_w^2} \quad ,$$

where m = the number of laboratories, and n = the number of data points per laboratory. For 3 laboratories and 5 data points per laboratory,

$$s_c = \sqrt{\frac{4}{3}s_b^2 + \frac{4}{5}s_w^2} \quad .$$

(6) Calculate the QC acceptance criteria for recovery in the OPR test by constructing a ± 2.6 s_c window around the average percent recovery ($\overline{\overline{X}}$, where 2.6 is the 97.5th percentile Student's *t*

value for 5 degrees of freedom (an estimated degrees of freedom based on the variance ratios observed with EPA Method 1625):

$$\text{Lower limit}(\%) = \overline{X} - 2.6s_c$$
$$\text{Upper limit}(\%) = \overline{X} + 2.6s_c$$

If more than 3 laboratories are used, the degrees of freedom for *t* will increase, but a complete calculation is beyond the scope of this document. An approximation of degrees of freedom equal to twice the number of laboratories will serve for most situations.

3.4.2.5 Matrix spike and matrix spike duplicate

Results of the MS/MSD analyses performed in the validation study are used to develop the MS/MSD QC acceptance criteria for Tier 2. Each laboratory will measure MS and MSD in each of the three samples. Calculate the MS and MSD performance criteria as follows.

(1) Calculate the mean and sample standard deviation of the recoveries of each MS/MSD pair, and then compute the overall mean recovery ($\overline{\overline{X}}$), the between-laboratory/matrix standard deviation of the 9 pairwise means (s_b), and the pooled within-laboratory/matrix standard deviation (s_w, as calculated in 3.4.2.2) for each target analyte.

(2) In order to allow for interlaboratory variability, calculate the combined standard deviation (s_c) for interlaboratory variability and estimation of the mean. For 3 laboratories and 3 matrices,

$$s_c = \sqrt{\frac{4}{3}s_b^2 + \frac{5}{6}s_w^2}$$

Derivation of the formula for other than 3 laboratories and 3 matrices is beyond the scope of this text.

(3) MS/MSD QC acceptance criteria for recovery - Calculate the QC acceptance criteria for recovery in the MS/MSD test by constructing a $\pm 2.2s_c$ window around the average percent recovery ($\overline{X}$) using the combined standard deviation. This factor comes from a *t* value for an estimated 7 degrees of freedom (based on this experimental design and variance ratios observed in Method 1625):

$$\text{Lower limit}(\%) = \overline{X} - 2.2s_c$$
$$\text{Upper limit}(\%) = \overline{X} + 2.2s_c$$

Note: For highly variable methods, it is possible that the lower limit for recovery will be a negative number. In these instances, the data should either be log-transformed and the recovery

window recalculated, or the lower limit established as "detected," as was done with some of the 40 CFR part 136, Appendix A methods.

(4) MS/MSD QC acceptance criteria for relative percent difference (RPD) - To evaluate a 95% confidence interval for precision, the RSD (computed using the pooled within laboratory standard deviation s_w of the MS/MSD samples divided by $\overline{\overline{X}}$) is multiplied by the square root of the 95th percentile F value with 1 degrees of freedom in the numerator and 3m degrees of freedom in the denominator multiplied by $\sqrt{2}$, where m is the number laboratories. The resulting multiplier on the RSD for 3 laboratories and 3 samples will then be 3.2. The QC acceptance criterion for precision in the MS/MSD test (RPD_{max}) is calculated as follows:

$$RPD_{max} = 3.2RSD.$$

3.4.2.6 Absolute and relative retention time

Establishing QC acceptance criteria for RT and RRT precision is problematic when multiple laboratories are involved because laboratories have a tendency to establish the chromatographic conditions that suit their needs. Calculating average RTs and RRTs based on different operating conditions will result in the establishment of erroneously wide windows. It is advised, therefore, that the organization developing the method specify to the participating laboratories the chromatographic conditions and columns to be used. Any future laboratories operating under different conditions will need to develop new acceptance criteria for RT and RRT precision.

Determine the average retention time, $\overline{RT}$, (or average relative retention time, $\overline{RRT}$), and the corresponding standard deviation (s) for each analyte and standard. Determine the upper and lower retention time (or relative retention time) limits using the following:

$$\text{Lower limit} = \overline{RT} - ts_{avg}\sqrt{1 + \frac{1}{n}}$$

$$\text{Upper limit} = \overline{RT} + ts_{avg}\sqrt{1 + \frac{1}{n}}$$

where the t value is the 97.5th percentile of a t distribution with n - 1 degrees of freedom and where n is the number of retention time or relative retention time data values to be used.

3.4.2.7 Blanks

Establish the QC acceptance criteria for blanks. The usual requirement is that the concentration of an analyte in a blank must be below the ML or below one-third (1/3) the regulatory compliance level, whichever is higher. In instances where the level of the blank is close to the regulatory compliance level or the level at which measurements are to be made, it may be necessary to require multiple blank measurements and establish the QC acceptance criteria based on the average of the blank measurements plus two standard deviations of the blank measurements.

3.4.2.8 Reference sample

Establish the QC acceptance criteria for the reference sample based on the error provided with the reference sample.

3.4.3 Quality Control Acceptance Criteria Development for New Methods at Tier 3

In Tier 3, a single sample collected from each of a minimum of nine industrial categories is analyzed in nine separate laboratories (one sample analyzed by each laboratory). Details for the characteristics and definitions of these samples are given in Chapter 4 of this guide. Because data gathered from nine laboratories lends itself to the statistical procedures used for interlaboratory method validation studies, the procedures suggested by ASTM and AOAC-International are particularly applicable and those procedures are preferred for development of QC acceptance criteria. However, QC acceptance criteria may also be developed for the Tier 3 methods in ways that are analogous to development of these criteria at Tiers 1 and 2, with minor modifications described below.

3.4.3.1 Method detection limits and minimum levels

Each laboratory participating in the validation must perform an MDL study as described in Section 3.4.1.1. The organization responsible for developing the new method must establish an MDL for the method, using a pooled MDL from the nine laboratories. A pooled MDL is calculated from m individual laboratory MDLs by computing the square root of the mean of the squares of the individual MDLs and multiplying the result by a ratio of *t*-values to adjust for the increased degrees of freedom.

$$MDL_{pooled} = \sqrt{\frac{d_1\left(\frac{MDL_{(Lab\,1)}}{t_{(0.99,d_1)}}\right)^2 + d_2\left(\frac{MDL_{(Lab\,2)}}{t_{(0.99,d_2)}}\right)^2 + \ldots d_m\left(\frac{MDL_{(Lab\,m)}}{t_{(0.99,d_m)}}\right)^2}{d_1 + d_2 + \ldots d_m}}\; t_{(0.99,d_1+d_2+\ldots d_m)} \quad ,$$

where m = the number of laboratories, and d_i = the number of replicates used by lab i to derive the MDL. In the case of 9 laboratories with 7 replicates per laboratory, the equation simplifies to:

$$MDL_{pooled} = \sqrt{\frac{MDL^2_{(Lab\,1)} + MDL^2_{(Lab\,2)} + \ldots MDL^2_{(Lab\,m)}}{9}}\;\frac{2.41}{3.14}$$

The organization responsible for developing the method must also use this MDL to develop an ML. Procedures for determining the ML are given in Section 3.4.1.1. Once established, the ML is used as the lowest calibration point. The instrument or analytical system is then calibrated at the ML and a minimum of two additional points to assess calibration linearity (Section 3.4.1.2) and to determine the number of calibration points required and how these points are spaced (Section 3.3.1).

3.4.3.2 Calibration linearity

Establish the RSD of the response factor, calibration factor or relative response based on the precision of the determinative technique, as described in Section 3.3.1. The RSD and the RSD limit for the response factor, calibration factor, or relative response is determined as follows:

(1) Calculate the average response factor ($\overline{\overline{RF}}$), average calibration factor ($\overline{\overline{CF}}$), or average relative response ($\overline{\overline{RR}}$) and pooled within-laboratory standard deviation (s_w) of the RF, CF, or RR determined for each analyte from each of the nine laboratories. The pooled standard deviation is computed as the square root of the mean of the squares of the sample standard deviations among the calibration results at each individual laboratory.

$$s_w = \sqrt{\frac{s^2_{(lab\ 1)} + s^2_{(lab\ 2)} + \ldots s^2_{(lab\ 9)}}{9}}$$

(2) Calculate the relative standard deviation (RSD) for each compound:

$$RSD = 100\frac{s_w}{\overline{\overline{RF}}}$$

(3) Calculate the maximum RSD for each analyte by the following:

$$RSD_{max} = kRSD,$$

where k is the square root of the 95th percentile of an F distribution with n - 1 degrees of freedom in the numerator and m(n - 1) degrees of freedom in the denominator, where m is the number of laboratories and n is the number of calibration points. For nine laboratories using a three-point calibration (n = 3), the value of k is 1.9, and for nine laboratories using a five-point calibration (n = 5), the value of k is 1.6.

Note: In the above equations, the $\overline{\overline{RF}}$ and RF terms should be replaced by $\overline{\overline{CF}}$ and CF or $\overline{\overline{RR}}$ and RR terms where appropriate.

3.4.3.3 Calibration verification

Using the average response factor or calibration factor from the initial calibration, calculate the upper and lower QC acceptance criteria for the calibration verification as follows:

(1) Determine "k" by multiplying the 97.5th percentile of a Student's *t* distribution with m(n-1) degrees of freedom times the square root of (1 + 1/mn), where there are n points in the calibration and m laboratories:

$$k = t\sqrt{(1 + \frac{1}{mn})}$$

For a three-point calibration with nine laboratories, the m(n - 1) Student's *t* value is 2.1 and for a five-point calibration, the Student's *t* value is 2.0, resulting in combined multipliers of 2.1 for both a three-point calibration and a five-point calibration.

Multiply k by the pooled standard deviation s_w found in Section 3.4.3.2.

(2) Calculate the upper and lower QC acceptance criteria for the response factors, calibration factors, or relative responses for each analyte by developing a window around the average response factor, calibration factor, or relative response by:

$$\text{Lower limit} = \overline{RF} - ks_w$$
$$\text{Upper limit} = \overline{RF} + ks_w$$

Note: In the above equations, the $\overline{RF}$ terms should be replaced by $\overline{CF}$ or $\overline{RR}$ terms where appropriate.

3.4.3.4 Initial and ongoing precision and recovery

For the IPR and OPR tests, QC acceptance criteria are calculated using the average percent recovery and the standard deviation of recovery from the IPR tests of four aliquots of the reference matrix and the OPR test of one aliquot of the reference matrix (for a total of five samples) in nine laboratories. The QC acceptance criteria are developed using the following steps:

(1) Calculate the average percent recovery ($\overline{X}$) for each analyte based on all data points from all laboratories, the between-laboratory standard deviation (s_b) of the mean results for each of the m laboratories (the standard deviation of the m laboratory averages $\overline{X}_{lab1}$, $\overline{X}_{lab2}$, ..., $\overline{X}_{labm}$), and the pooled within-laboratory standard deviation (s_w) of the five samples calculated as in 3.4.3.2. Note: the organization responsible for developing the method must ensure that all laboratories are spiking IPR and OPR samples at the same concentration.

(2) IPR QC acceptance criteria for precision - To calculate a 95% confidence interval for precision, the RSD (computed as s_w divided by $\overline{X}$) is multiplied by the square root of the 95th percentile F value with 3 degrees of freedom in the numerator and 4m degrees of freedom in the denominator. The resulting multiplier for nine laboratories will be 1.7. The QC acceptance criterion for precision in the IPR test (RSD_{max}) for 9 laboratories is calculated as follows:

$$RSD_{max} = 1.7RSD$$

(3) IPR QC acceptance criteria for recovery -Calculate the combined standard deviation for interlaboratory variability and estimation of the mean (s_c) as:

$$s_c = \sqrt{(1+\frac{1}{m})s_b^2 + (\frac{1}{4}-\frac{1}{n})s_w^2} \quad ,$$

where m = the number of laboratories, and n = the number of data points per laboratory. For 9 laboratories and 5 data points per laboratory,

$$s_c = \sqrt{\frac{10}{9}s_b^2 + \frac{1}{20}s_w^2} \quad .$$

(4) Calculate the QC acceptance criteria for recovery in the IPR test by constructing a ± 2.3 s_c window around the average percent recovery ($\overline{\overline{X}}$, where 2.3 is the 97.5th percentile Student's *t* value for 10 degrees of freedom (an estimated degrees of freedom based on the variance ratios observed with EPA Method 1625):

$$\text{Lower limit}(\%) = \overline{\overline{X}} - 2.3s_c$$
$$\text{Upper limit}(\%) = \overline{\overline{X}} + 2.3s_c$$

If more than 9 laboratories are used, the degrees of freedom for t will increase, but a complete calculation is beyond the scope of this document. An approximation of degrees of freedom equal to the number of laboratories will serve for most situations.

(5) OPR QC acceptance criteria for recovery - Calculate the combined standard deviation for interlaboratory variability and estimation of the mean (s_c) as:

$$s_c = \sqrt{(1+\frac{1}{m})s_b^2 + (1-\frac{1}{n})s_w^2} \quad ,$$

where m = the number of laboratories, and n = the number of data points per laboratory. For 9 laboratories and 5 data points per laboratory,

$$s_c = \sqrt{\frac{10}{9}s_b^2 + \frac{4}{5}s_w^2}$$

(6) Calculate the QC acceptance criteria for recovery in the OPR test by constructing a ± 2.1 s_c window around the average percent recovery ($\overline{\overline{X}}$, where 2.1 is the 97.5th percentile Student's *t*

value for 19 degrees of freedom (an estimated degrees of freedom based on the variance ratios observed with EPA Method 1625):

$$\text{Lower limit}(\%) = \overline{X} - 2.1 s_c$$
$$\text{Upper limit}(\%) = \overline{X} + 2.1 s_c$$

If more than 9 laboratories are used, the degrees of freedom for *t* will increase, but a complete calculation is beyond the scope of this document. An approximation of degrees of freedom equal to twice the number of laboratories will serve for most situations.

3.4.3.5 Matrix spike and matrix spike duplicate

Results of the MS/MSD analyses performed in the Tier 3 validation study are used to develop the MS/MSD QC acceptance criteria for Tier 3. Calculate the MS and MSD performance criteria as follows.

(1) Calculate the percent recovery ($\overline{X}$) and the between-laboratory standard deviation (s_b) of the mean results for each of the nine laboratories and also the pooled within-laboratory standard deviation (s_w as calculated as in 3.4.3.2) for each target analyte using the MS and MSD analyses.

(2) In order to allow for interlaboratory variability, calculate the combined standard deviation (s_c) for interlaboratory variability and estimation of the mean as:

$$s_c = \sqrt{(1 + \frac{1}{m}) s_b^2 + \frac{1}{2} s_w^2}.$$

where m = the number of laboratories. For nine labs,

$$s_c = \sqrt{\frac{10}{9} s_b^2 + \frac{1}{2} s_w^2}$$

(3) MS/MSD QC acceptance criteria for recovery - Calculate the QC acceptance criteria for recovery in the MS/MSD test by constructing a ± 2.2 s_c window around the average percent recovery $\overline{X}$ using the combined standard deviation. This factor comes from a t value for an estimated 11 degrees of freedom (based on this experimental design and variance ratios observed in Method 1625):

$$\text{Lower limit}(\%) = \overline{X} - 2.2 s_c$$
$$\text{Upper limit}(\%) = \overline{X} + 2.2 s_c$$

Note: For highly variable methods, it is possible that the lower limit for recovery will be a negative number. In these instances, the data should either be log-transformed and the recovery window recalculated, or the lower limit established as "detected," as was done with some of the 40 CFR part 136, Appendix A methods.

(4) MS/MSD QC acceptance criterion for relative percent difference (RPD) - To calculate a 95% confidence interval for precision, the RSD (computed using the pooled within-laboratory standard deviation, s_w, of the MS/MSD samples divided by $\overline{\overline{X}}$) is multiplied by the square root of the 95% percentile F value with 1 degree of freedom in the numerator and m degrees of freedom in the denominator multiplied by $\sqrt{2}$. The resulting multiplier on the RSD for nine laboratories will be 3.2. The QC acceptance criterion for precision in the MS/MSD test (RPD_{max}) is calculated as follows:

$$RPD_{max} = 3.2RSD.$$

3.4.3.6 Absolute and relative retention time

Establishing QC acceptance criteria for RT and RRT precision is problematic when multiple laboratories are involved because laboratories have a tendency to establish the chromatographic conditions that suit their needs. Calculating average RTs and RRTs based on different operating conditions will result in the establishment of erroneously wide windows. It is advised, therefore, that the organization developing the method specify to the participating laboratories the chromatagraphic conditions and columns to be used. Any future laboratories operating under different conditions will need to develop new acceptance criteria for RT and RRT precision.

(1) Using replicate RT and/or RRT data, calculate the upper and lower QC acceptance criteria for each analyte using the procedures in the calibration verification test in Section 3.4.1.3.

(2) Determine the average retention time, $\overline{RT}$ (or average relative retention time, $\overline{RRT}$), and the corresponding standard deviation (s) for each analyte and standard. Determine the upper and lower retention time (or relative retention time) limits using the following:

$$\text{Lower limit} = \overline{RT} - ts\sqrt{1 + \frac{1}{n}}$$

$$\text{Upper limit} = \overline{RT} + ts\sqrt{1 + \frac{1}{n}}$$

where the t value is the 97.5th percentile of a t distribution with n - 1 degrees of freedom, where n is the number of retention time or relative retention time data values to be used.

3.4.3.7 Blanks

Establish the QC acceptance criteria for blanks. The usual requirement is that the concentration of an analyte in a blank must be below the ML or below one-third (1/3) the regulatory compliance level, whichever is higher. In instances where the level of the blank is close to the regulatory compliance level or the level at which measurements are to be made, it may be necessary to require

multiple blank measurements and establish the QC acceptance criteria based on the average of the blank measurements plus two standard deviations of the blank measurements.

3.4.3.8 Reference sample

Establish the QC acceptance criteria for the reference sample based on the error provided with the reference sample.